EXHIBIT 1

~~FILED UNDER SEAL~~

PUBLIC REDACTED VERSION

	A	B	C	D	E	F	G	H
1		The Times's Proposed Search Terms	Total Documents Hit by Term Including Families		OpenAI's Initial Search Term Proposal	Total Documents Hit by Term Including Families		Court's Ruling
2	**Unique, De-duplicated Hit Counts for all Terms Combined**		**546,902**			**131,601**		
3								
4	**Section A: Terms Relating to the Products' Outputs**							
5								
6	1	paywall	861		N/A			
7	2	antiregurg*	525		N/A			
8	3	regurg*	86		regurg* AND (ChatGPT OR GPT-3.5 OR GPT-3.5T OR "GPT-3.5 turbo" OR GPT-4 OR GPT-4T OR "GPT-4 turbo") AND copyright*	431		
9	4	(Bing* w/5 index) OR (search w/5 index) OR (search w/5 result*) OR (ranking* w/5 engine) OR ((knowledge OR entity) w/5 graph) OR ((prompt* OR query) w/5 logs) OR (user w/5 (queries OR prompt*))	47,148		N/A			
10	5	Retriev* w/15 ("based model" OR knowledge OR document* OR information)	13,679		N/A			
11	6	(retriev* w/5 augmen*) OR RAG	69		"Retrieval Augmented Generation"	169		
12	7				RAG	1,301		
13	8	(generat* OR ground* OR synthetic OR semantic OR deriv*) w/25 search*	16,979		N/A			
14	9	[redacted]	5,277		[redacted] AND (ChatGPT OR GPT-3.5 OR GPT-3.5T OR "GPT-3.5 turbo" OR GPT-4 OR GPT-4T OR "GPT-4 turbo")	1,147		
15								
16	**Section B: Terms Related to GenAI and Journalism**							
17								
18	11	"NY Times" OR NYT OR "New York Times" OR "The Times" OR *nytimes* OR "The Athletic" OR *theathletic* OR *wirecutter*	11,572		(ChatGPT OR GPT-3.5 OR GPT-3.5T OR "GPT-3.5 turbo" OR GPT-4 OR GPT-4T OR "GPT-4 turbo") w/25 ("New York Times" OR NYTimes OR NYT OR "NY Times")	1,169		
19	12	(ethic* w/5 committee) OR (ethic* w/5 audit) OR (ethic* w/5 review) OR "ethic* Team" OR (red w/5 team*) OR "AI impact assessment" OR Aether OR "pre-deployment reviews" OR "stress testing" OR "responsible AI" OR RAI OR "Python Risk Identification Tool" OR PyRIT OR (compliance w/5 audit) OR (compliance w/5 committee)	28,300		N/A			
20	13	(Traffic OR search* OR clickthrough OR click-through OR "click through" OR CTR) w/25 (*GPT* OR dv3 OR dv-3 OR davinci-3 OR genai OR generative OR Bing OR Copilot OR Azure OR genai OR SLM OR Microsoft OR Msft OR Prometheus OR Orchestrat*)	33,819		N/A			
21								
22	**Section C: Terms Relating to The Times's Legal Claims**							
23	14	(Concern* OR importan* OR valu* OR need* OR neces* OR *legal* OR *law* OR complain* OR claim* OR infring* OR violat* OR right* OR "intellectual property" OR IP) w/50 (*gpt* OR dv3 OR generative OR genai OR SLM OR Bing OR Copilot OR Azure OR Microsoft OR Msft OR Prometheus OR Orchestrat*)	147,731		N/A			
24	15	Copyright* OR &#169 OR &copy OR DRM OR "digital rights management" OR (fair w/5 use)	25,747		copyright* w/5 (search* OR clearance*)	63,863		
25	16	(Trademark* OR TM OR Lanham OR dilut* OR tarnish* OR &#8482 OR &trade OR &#174 OR &reg) w/25 (*gpt* OR dv3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai OR Microsoft OR Msft)	1,437		N/A			
26	17	Copy* w/5 Shield	3		N/A			
27	18	Indemni*	3,226		N/A			
28	19	((IP OR "intellectual property" OR copy*) AND (polic* OR procedure* OR practice* OR conduct* OR action* OR guidance* OR guideline* OR SOP OR "standing operating procedure")) w/25 (*gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai OR Microsoft OR Msft OR transformer OR "neural network*" OR "machine learning" OR Prometheus OR Orchestrat*)	1,045		N/A			
29	20	CMI OR DMCA OR 1202 OR watermark* OR (copyright w/5 management)	19,492		N/A			
30	21	data w/25 filter*	18,890		(filter* OR mask*) w/10 (train* w/10 (ChatGPT OR GPT-3.5 OR GPT-3.5T OR "GPT-3.5 turbo" OR GPT-4 OR GPT-4T OR "GPT-4 turbo"))	73		
31	22	(misuse w/3 detect*) OR (domain w/3 filter) OR (action w/3 monitor*) OR (us* w/3 polic*)	16,258		N/A			
32	23	(sketchy OR *reliab* OR *trust*) w/15 (data OR content* OR safety OR polic*)	24,378		N/A			
33	24	(content OR safety OR policy) w/25 (filter* OR mask*)	17,085		N/A			
34								
35	**Section D: Terms Relating to Model Training**							
36	25	(*train* OR "fine tun*" OR fine-tun* OR finetun* OR instruct* OR *param* OR corpus OR corpora OR token* OR RLHF OR "proximal policy optimization" OR "data set" OR dataset* OR "test data") w/25 (discriminator OR curricul* OR classif* OR *score* OR frequen* OR contamination* OR duplicat* OR regression* OR cluster* OR feature* OR label* OR attribute* OR categ* OR value* OR variable* OR inference OR distillation OR recipe*)	221,351		book* w/5 (source* OR train* OR data* OR collection)	21,541		
37	26	(*train* OR "fine tun*" OR fine-tun* OR finetun* OR instruct* OR *param* OR corpus OR corpora OR token* OR RLHF OR "proximal policy optimization" OR "data set" OR dataset* OR "test data") w/25 (rule* OR guidance* OR weight* OR perplexity OR ROUGE OR F1 OR classif* OR benchmark OR whitelist OR blacklist OR safeguard* OR guardrail* OR quality OR recitation* OR scrap* OR checkpoint)	130,910		See row above			

	A	B	C	D	E	F	G	H
1		The Times's Proposed Search Terms	Total Documents Hit by Term Including Families		OpenAI's Initial Search Term Proposal	Total Documents Hit by Term Including Families		Court's Ruling
2	**Unique, De-duplicated Hit Counts for all Terms Combined**		**546,902**			**131,601**		
38	27	data w/10 (clean* OR *process* OR curat* OR select* OR augment* OR sampl* OR gather* OR copy* OR publish*)	71,644		N/A			
39	28	(CC NOT w/25 (.com OR (To AND BCC))) w/50 (data OR content OR *train* OR dataset* OR *gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai OR Prometheus OR Orchestrat*)	123,352		N/A			
40	29	(train* OR stage* OR layer*) w/5 (learn* OR iteration OR cycle OR batch OR epoch)	30,177		N/A			
41	30	C4	13,205		N/A			
42	31	webtext* OR WT OR WT2	5,792		webtext	7,087		
43	32				commoncrawl OR "common crawl"	7,669		
44	33	(*gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai OR Prometheus OR Orchestrat*) w/50 (journali* OR news* OR report* OR info* OR knowledge OR article* OR story OR stories OR publish* OR headline* OR cover* OR syndicat* OR podcast* OR lede OR scoop* OR editorial OR opinion)	78,357		(ChatGPT OR GPT-3.5 OR GPT-3.5T OR "GPT-3.5 turbo" OR GPT-4 OR GPT-4T OR "GPT-4 turbo") w/25 (journali* OR news*)	4,255		
45	34	(impact* w/5 assess*) w/25 (*gpt* OR dv3 OR dv-3 OR davinci-3 OR genai OR generative OR SLM OR Bing OR Copilot OR Azure)	44		N/A			
46	35	(*gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai) w/5 metric*	2,512		N/A			
47								
48	**Section E: Terms Relating to Technical Documentation and Processes**							
49	36	(reward OR policy) w/10 model*	18,701		N/A			
50	37	(few OR one OR zero) w/10 shot*	27,764		N/A			
51	38	(training OR stag* OR layered) w/5 (learn* OR iteration OR cycle OR batch OR epoch OR curve)	21,245		N/A			
52	39	"model completion"	705		N/A			
53	40	data w/25 scrap*	5,539		N/A			
54	41	("validation data" OR "The Pile" OR preprocessing OR "test set" OR "training set" OR "evaluation set" OR "text extraction" OR "semantic analysis" OR "source attribution") w/25 (discriminator OR curricul* OR classif* OR *score* OR frequen* OR contamination* OR duplicat* OR regression* OR cluster* OR feature* OR label* OR attribute* OR categ* OR value* OR variable* OR inference OR distillation OR recipe*)	3,756		N/A			
55	42	("validation data" OR "The Pile" OR preprocessing OR "test set" OR "training set" OR "evaluation set" OR "text extraction" OR "semantic analysis" OR "source attribution") w/25 (rule* OR guidance* OR weight* OR perplexity OR ROUGE OR F1 OR classif* OR benchmark OR whitelist OR blacklist OR safeguard* OR guardrail* OR quality OR recitation* OR scrap* OR checkpoint)	1,676		N/A			
56	43	KPI OR PRD OR (performance w/25 indicator)	73		N/A			
57	44	KPI OR PRD OR (performance w/25 indicator)	73		N/A			
58	45	measur* w/5 pipe*	130		N/A			
59	46	(model w/25 card) OR "product brief" OR "product requirement" OR "data card" OR datasheet OR "data doc*"	430		N/A			
60	47	(product* OR design* OR technical OR architecture OR system* OR launch*) w/10 (specification* OR requirement* OR brief* OR report* OR chart OR log)	42,250		N/A			
61	48	embed* w/15 (function OR space OR vector OR index* OR indices)	9,855		N/A			
62	49	relevance OR "query tuning" OR (source* w/5 (trust* OR credib*)) OR (standard* w/5 editor*) OR (criteri* w/5 (exclusion OR inclusion))	15,904		N/A			
63	50	[redacted]	1,299		N/A			
64	51	Azure AND compute*	20,934		Azure AND ("super computer" OR supercomputer)	1,335		
65	52	(journali* OR news* OR report* OR info* OR knowledge OR article* OR story OR stories OR publish* OR headline* OR cover* OR syndicat*) AND ((data* w/5 quality) OR (high w/5 quality))	34,433		dataset* w/10 "high quality"	14,570		
66								
67	**Section F: Terms Relating to Colloboration with Microsoft**							
68								
69	53	[redacted]	13,675		N/A			
70	54	[redacted]	834		N/A			
71	55	crescendo	167		N/A			
72	56	mail(*@microsoft.com) AND (data* OR train* OR *gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR Genai OR SLM OR Bing OR Copilot OR Azure OR *compute*)	13,645		(Microsoft OR MSFT) AND ("super computer" OR supercomputer)	2,675		
73	57				mail(*@microsoft.com) AND (train* w/10 (data* OR material* OR source*))	151		
74	58	MS w/25 (data* OR train*)	3,719		N/A			
75	**Section G: Search Terms Related to the Market for and Value of Content**							
76	59	(*Licens* OR agreement* OR compensat* OR pay OR paid) AND (*gpt* OR dv3 OR dv-3 OR davinci-3 OR genai OR generative OR SLM OR Bing OR Copilot OR Azure OR Prometheus OR Orchestrat*)	46,947		N/A			
77	60	(license OR agreement) AND (train* w/15 model*)	6,127		N/A			
78	61	(public* OR "open source" OR "creative commons" OR government OR 1920 OR 1921 OR 1922 OR 1923 OR 1924 OR 1925 OR 1926 OR 1927 OR 1928) w/25 (*gpt* OR dv3 OR genai OR generative OR SLM OR Bing OR Copilot OR Azure)	16,071		N/A			

	A	B	C	D	E	F	G	H
1		The Times's Proposed Search Terms	Total Documents Hit by Term Including Families		OpenAI's Initial Search Term Proposal	Total Documents Hit by Term Including Families		Court's Ruling
2	**Unique, De-duplicated Hit Counts for all Terms Combined**		**546,902**			**131,601**		
79	62	(MAI OR MAI-1 OR Phi OR Phi-2 OR Phi-3) w/50 (azure OR CPU OR GPU OR compute* OR invest* OR financ* OR profit* OR revenue* OR stake OR money OR process* OR server* OR hardware)	515		N/A			
80	63				(MAI OR MAI-1 OR Phi OR Phi-2 OR Phi-3) AND azure	699		
81								
82	**Section H: Search Terms OpenAI Seemingly Re-Used from Other Cases**							
83								
84	64				Libgen* OR "Library Genesis"	4,748		
85	65				"books corpus" OR (book* w/5 corp*) OR bookscorp*	2,721		
86	66				Gutenberg	3,111		
87	67				z-library OR zlibrary OR zlib	1,287		
88	68				"Google books"	1,698		
89	69				B-ok	1,077		
90	70				"Open Library"	1,520		
91	71				book* w/5 ("data quality" OR "high quality")	174		
92	72				smashwords	1,188		
93	73				Bibliotik	18		
94	74				"Anna's Archive"	11		
95					Sci-Hub OR scihub OR "sci hub"	342		